BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 9: e183060
doi: 10.3897/biss.9.183060

OPEN ACCESS

Conference Abstract

Wikibase to the Rescue? Exploring the Use of Wikibase to Extend the Sherwin Carlquist Digital Extended Specimen Network

Jason H. Best‡, Sam Ekberg‡, Mare Nazaire§, Ana Niño‡, Krishna Shenoy‡

‡ Botanical Research Institute of Texas, Fort Worth, United States of America
§ California Botanic Garden, Claremont, United States of America

Corresponding author: Jason H. Best (jbest@brit.org)

Received: 19 Dec 2025 | Published: 23 Dec 2025

Citation: Best JH, Ekberg S, Nazaire M, Niño A, Shenoy K (2025) Wikibase to the Rescue? Exploring the Use of Wikibase to Extend the Sherwin Carlquist Digital Extended Specimen Network. Biodiversity Information Science and Standards 9: e183060. https://doi.org/10.3897/biss.9.183060

Abstract

Introduction

The Extended Specimen Network (ESN) (Lendemer et al. 2019) concept recognizes that scientific value extends beyond physical specimens to include field notes, field photographs, publications, and other research products. The Sherwin Carlquist ESN (Anonymous 2025) unites the legacy of Dr. Sherwin Carlquist (1930–2021), a prominent American botanist whose career focused on island biogeography and plant anatomy. This ESN connects Carlquist's physical specimens at California Botanic Garden (CalBG) with his archival materials at the Botanical Research Institute of Texas (BRIT), demonstrating both the potential and challenges of creating comprehensive ESNs across multiple institutions and digital platforms.

Project Infrastructure and Goals

The project established several objectives:

- imaging and transcribing approximately 9,000 specimens
- imaging and transcribing 26 field notebooks;
- digitizing and describing approximately 100,000 field photographs with metadata

© Best J et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

- establishing linkages between field images and botanical specimens
- applying discovered linkages across digital platforms

Digital records are distributed across the Consortium of California Herbaria (CCH2)*[1] , a Symbiota*[2] portal for biological specimens, and the Portal to Texas History (PTH)*[3] for archival materials.

Cross-Platform Challenges

Implementation of the ESN on the digital platforms revealed some limitations. The PTH's restriction to Dublin Core metadata*[4] forced use of generic relationship terms "references" and "isReferencedBy"—both lacking semantic specificity. PTH also offered limited external linking capability (with relationships and URLs only describable in free text) and lacked the ability to search or filter to target only archival records with linked resources. Symbiota initially lacked functionality for linking to external archival items, a limitation addressed by creating a linked resources module through collaboration with the Symbiota developers.

Both platforms shared limitations: difficulty creating complex linkages for entities like people, taxa, and publications; limited integration with existing linked data networks; and insufficient semantic richness for expressing some ESN relationships.

Linked Data Approach

Linked data represents structured information interlinked through semantic relationships, enabling machine-readable queries. This aligns naturally with ESNs, where specimens, images, notes, and publications form interconnected information networks. Wikidata, containing millions of structured items, represents the largest openly accessible linked data resource. Wikibase, the open-source software underlying Wikidata, offers institutions the opportunity to implement similar infrastructure while connecting to the broader linked open data ecosystem. We established a Wikibase instance*[5] to experiment with and explore the capabilities and potential of the platform as a complementary extension of the Carlquist ESN created in our primary digital platforms.

Implementation Methods

To populate the Wikibase ESN, we began by aggregating data using multiple approaches: Sickle*[6] (a Python module) retrieved PTH archival records; Symbiota's Darwin Core Archive (GBIF 2021) provided specimen data; and scripts using the Biodiversity Heritage Library (BHL) API*[7] retrieved publication data. Data preparation involved iterative cleaning and standardization using OpenRefine*[8]. Wikibase properties were created using the WikibaseIntegrator*[9] Python module, datasets were uploaded via WikibaseIntegrator with SPARQL*[10] queries, and relationships established using the Wikibase QuickStatements tool*[11].

Results

The Carlquist ESN Wikibase instance contains 8,941 specimen records linking to California Botanic Garden/CCH2; 13,462 field image and notebook records linking to BRIT/PTH; 301 publication records (37 linking to BHL); and biographical data for Carlquist linking to Wikipedia and Wikidata. Each record maintains bidirectional links to source platforms, creating cross-institutional relationship networks.

This integrated approach enables capabilities that were challenging or impossible within original platforms: specimens link to field images across institutions; publications connect to referenced specimens and images; complex queries retrieve related items across data types; and the ESN connects to broader linked data networks through Wikidata integration.

Conclusion

Wikibase offers significant promise as an ESN enhancement layer. Rather than replacing existing infrastructure, it provides complementary capabilities: specimen and extension relationships across platforms; expansion beyond platform limitations; comprehensive complex relationships; cross-dataset queries; inclusion of entities not in collection systems; and openness for community contribution. Before Wikibase can be fully embraced for implementation of an ESN, more effort will be required to accurately model entities such as specimens, field notebooks, and field photos. Nevertheless, this experimental implementation demonstrates the feasibility and value for institutions seeking to extend digital collections beyond traditional platform constraints while participating in the linked open data ecosystem.

Keywords

Extended Specimen Network, Wikidata, Linked Open Data, natural history collections, archives

Presenting author

Jason H. Best

Conflicts of interest

The authors have declared that no competing interests exist.

References

- Anonymous (2025) The Sherwin Carlquist Collection. https://fwbg.org/science-conservation/brit-library/the-sherwin-carlquist-collection/. Accessed on: 2025-11-26.

- GBIF (2021) Darwin Core Archives – How-to Guide, version 2.2. GBIF Secretariat URL: https://ipt.gbif.org/manual/en/ipt/3.0/dwca-guide
- Lendemer J, Thiers B, Monfils AK, Zaspel J, Ellwood ER, Bentley A, LeVan K, Bates J, Jennings D, Contreras D, Lagomarsino L, Mabee P, Ford LS, Guralnick R, Gropp RE, Revelez M, Cobb N, Seltmann K, Aime MC (2019) The Extended Specimen Network: A Strategy to Enhance US Biodiversity Collections, Promote Research and Education. BioScience 70 (1): 23-30. https://doi.org/10.1093/biosci/biz140

Endnotes

*1 Consortium of California Herbaria data portal (CCH2) - https://cch2.org/

*2 Symbiota, a natural history collections managment platform - https://symbiota.org/

*3 Portal to Texas History - https://texashistory.unt.edu/

*4 Dublin Core metadata, a specification for describing general resources - https://www.dublincore.org/specifications/dublin-core/

*5 Sherwin Carlquist experimental Extended Specimen Network using Wikibase software - https://extended-specimen.wikibase.cloud/

*6 Sickle, a Python module that serves as an OAI-PMH client - https://github.com/mloesch/sickle

*7 Biodiversity Heritage Library API - https://about.biodiversitylibrary.org/tools-and-services/developer-and-data-tools/#APIs

*8 OpenRefine, an open source tool for cleaning and transforming data - https://openrefine.org/

*9 WikibaseIntegrator, a Python module for manipulating Wikibase data - https://github.com/LeMyst/WikibaseIntegrator

*10 SPARQL, a query language used to retrieve and manipulate data stored in RDF format - https://www.w3.org/TR/sparql11-query/

*11 QuickStatements, a tool for editing Wikidata and Wikibase items - https://www.wikidata.org/wiki/Help:QuickStatements